

Funkcjonalistyczna teoria umysłu, przetwarzanie informacji a myślenie ludzkie*

(Functionalist theory of mind,
processing of information and human thinking)

TOMASZ HUZAREK

Uniwersytet Mikołaja Kopernika w Toruniu

tomhuzar@gmail.com

Streszczenie. Funkcjonalizm jako teoria umysłu zaproponowana przez Hilarego Putnama traktuje stany umysłu jako stany funkcjonalne, a własności umysłu jako własności funkcjonalne. Według tej teorii stany mentalne mogą być realizowane przez różne stany fizyczne. W ten sposób funkcjonalizm stał się filozoficzną podstawą obliczeniowych, komputacyjnych teorii umysłu, wedle których myślenie ludzkie polega na obliczeniach, a świadome doznania powstają na skutek procesu obliczeniowego. Główny problem, jaki podejmuję w artykule, wyraża pytanie: czy rzeczywiście myślenie ludzkie można zredukować do procesu obliczeniowego o charakterze algorytmicznym? Eksperymenty myślowe J. Searle'a czy N. Blocka, jak również nasze potoczne doświadczenie wskazują, iż nasze myślenie jest czymś więcej niż tylko algorytmicznym przetwarzaniem informacji.

Słowa kluczowe: funkcjonalizm; umysł; przetwarzanie informacji; samoświadomość; intencjonalność

Abstract. Functionalism as a theory of mind was proposed by Hilary Putnam. It treats the states of mind as functional states and properties of mind as functional properties.

* W roku 2006 w reaktywowanym wtedy czasopiśmie Koła Filozoficznego Studentów KUL „Filozofia” (VI 2006) ukazał się niewielki tekst zatytułowany *Człowiek a maszyna*. Niniejszy artykuł stanowi jego przeformułowanie i poszerzenie.

According to this theory mental states can be accomplished by various physical states. This is how functionalism became the philosophical base for computational theories of mind, which say that human thinking is just calculation and aware experiences come into existence as a result of computational process. The main theme of my article is the question: Can human thinking be reduced to algorithmical computational process? Mental experiment which was taken by J. Searle and N. Block, likewise our common experience shows that our thinking is something more than just algorithmical repetition of information.

Key words: functionalism; mind; processing of information; self-awareness; intentionality

1. Funkcjonalizm w teorii umysłu

Matthew Nagle, dwudziestopięcioletni Amerykanin, stał się pierwszym człowiekiem na świecie, któremu wszczepiono do mózgu układ elektrod odczytujący polecenia. Cztery lata wcześniej Matthew został zraniony nożem, wskutek czego naruszony został centralny układ nerwowy. Ostatecznie nastąpił paraliż ciała od szyi w dół. Zabieg, przeprowadzony przez specjalistów z New England Sinai Hospital, polegał na wprowadzeniu do mózgu Nagle'a około stu elektrod o grubości dziesiątych części milimetra w ten obszar kory mózgowej, który jest odpowiedzialny za ruchy ciała. Teraz Nagle może za pomocą myśli, to jest, zdaniem owych specjalistów, za pomocą odpowiednich serii wyładowań w neuronach, sterować urządzeniami technicznymi. Sygnały z mózgu są przekazywane do komputera, który tłumaczy je na rozkazy dla obsługiwanego sprzętu. Wystarczy myśłami przeciągać kursor po ekranie. W ten sposób można na przykład włączyć i wyłączyć telewizor, zmieniać kanały i regulować głośność. W zasadzie można sterować dowolnym urządzeniem, jeśli ma odpowiedni interfejs. Zajmujący się przypadkiem Nagle'a specjaliści są przekonani, że to rewolucyjne dokonanie wskazuje, iż w przyszłości będzie można przeprowadzić zabiegi pozwalające odzyskać całkowitą władzę w sparaliżowanych kończynach. Aparat wielkości telefonu komórkowego w odpowiedzi na sygnały z mózgu sterowałby skurczami mięśni za pomocą impulsów elektrycznych. Trudność leży w tym, że nawet za mało skomplikowany ruch odpowiedzialnych jest wiele partii mięśni. Droga jest więc jeszcze daleka.

To niebywałe osiągnięcie medycyny może być rozumiane jako potwierdzenie niektórych konsekwencji wyprowadzanych z powszechnie przyjmowanych naturalistycznych koncepcji filozoficznych. Stanowią one alternatywę wobec substancjalnych teorii w filozofii umysłu. Jedno z takich alternatywnych stanowisk zaprezentował Hilary Putnam.

Funkcjonalizm Putnama zaliczany jest do stanowisk naturalistycznych w filozofii umysłu. Współczesny naturalizm jest niejednorodny, jednak wspólną cechą filozofii tego nurtu jest przekonanie, że jedyna rzetelna wiedza na temat umysłu może być osiągnięta w ramach nauk ścisłych, które przewyższą potoczne i filozoficzne mity, które narosły przez lata wokół zagadnienia umysłu. Zadaniem zaś filozofii „pozostaje systematyzacja szczególnie doniosłych wyników wybranych teorii naukowych lub analiza ich metodologicznych i ontologicznych założeń” (Miłkowski, Poczo but 2012, 12). Redaktorzy *Przewodnika po filozofii umysłu* wskazują na wiele odmian naturalizmu w dziedzinie filozofii umysłu: naturalizm filozoficzny (D. Papineau) i naukowy (W.V.O. Quine), naturalizm biologiczny (J. Searle), naturalizm reprezentacyjny oparty na gruncie teorii informacji (F. Dretske), naturalizm ewolucyjny (R. Dawkins, D. Dennett), naturalizm eliminacyjny, naturalizm redukcyjny (J. Bickle, J. Kim), naturalizm nie-redukcyjny (D. Davidson, H. Putnam, R. Van Gulick) oraz naturalistyczny dualizm własności (D. Chalmes).

Pisma Hilarego Putnama odegrały zasadniczą rolę w ukształtowaniu się dominujących obecnie koncepcji umysłu. Funkcjonalizm, chociaż zrobił błyskawiczną karierę w filozofii umysłu, stanowi dzisiaj bardzo szeroki i zróżnicowany nurt. Stanowisko funkcjonalizmu zaprezentował Putnam w artykule *Umysły i maszyny* (1960). Ta propozycja miała stanowić alternatywę wobec substancjalnych teorii w filozofii umysłu, którym Putnam zarzucał błąd kategorialny, i być teorią legitymizującą obliczeniowe podejście do umysłu. Po opublikowaniu przez Putnama *Psychological Predicates* (1967) niemal zgodnie uznano, że teorię identyczności oraz behawioryzm psychologiczny należy porzucić, nie są bowiem adekwatne w opisie tego, czym jest umysł (Hensel 2011). Hensel sugeruje, że pojawienie się funkcjonalizmu Putnama poprzedzało dialektyczne napięcie dwóch teorii w fi-

lozofii umysłu: z jednej strony behawioryzmu logicznego – opartego na solidnej metodologii oraz przekonującej koncepcji uczenia się języka, ale jednocześnie nieintuicyjnego, sugerującego bowiem, że stany mentalne (ból, gniewu itd.) nie istnieją; z drugiej zaś strony redukcjonizm psychofizyczny, który nie podważa faktu istnienia stanów mentalnych, ujmując je jako stany mózgu, ale powątpiewano w jego moc wyjaśniającą.

Teoria funkcjonalizmu Putnama ewoluowała. Wczesne stanowisko, które Hensel nazywa „słabym funkcjonalizmem”, to pogląd sceptyczny, który „z jednej strony wykorzystywał intuicję, że stany umysłu mają charakter relacyjny, do obalenia behawioryzmu logicznego i fizykalizmu typów, z drugiej zaś wskazywał na arbitralność wszelkich rozstrzygnięć w kwestii umysł–ciało” (Hensel 2011, 31). Później jednak Putnam przekształca tezy funkcjonalizmu słabego w doktrynę pozytywną odrzucającą tezę o arbitralności w filozofii umysłu, by ostatecznie w ogóle odejść od teorii funkcjonalizmu (Bremer 2010). Około połowy lat 70. autor pracy *Znaczenie wyrazu „znaczenie”* (1975) odwraca się od teorii funkcjonalizmu z powodu trudności pogodzenia tej teorii umysłu z doktryną, wedle której stany mentalne zawdzięczają swój wewnętrzny charakter otoczeniu bądź środowisku podmiotów¹. Późniejsze jednak koncepcje Putnama nie miały już takiego oddziaływania jak zaproponowany przez niego funkcjonalizm, który w licznych odmianach jest nadal dominujący w filozofii umysłu.

Nie wchodząc w szczegóły rozwoju koncepcji Putnama, można powiedzieć, że jego stanowisko – zanim zaczął porzucać funkcjonalizm – przyjmuje *zasadę izomorfizmu psychologicznego*, wedle którego „dla dowolnej teorii psychologicznej lub epistemologicznej może istnieć niestandardowy model (np. robot), bowiem własności psychiczne mają naturę funkcjonalną” (Hensel 2011, 39). Oznacza to, że istotą własności umysłowych jest funkcja, a nie materialne podłoże, które pozwala tę funkcję realizować.

¹ John Heil nazywa tę drugą doktrynę „eksternalizmem”: pewne stany umysły – te, które mają „zawartość intencjonalną” – zależą od czegoś więcej niż tylko od wewnętrznego fizycznego charakteru podmiotów, którym są przypisywane, zob. (Heil 2001, 129–130).

Funkcjonalistyczna teoria umysłu nawiązuje w sposób krytyczny do idei behawiorystycznych oraz teorii identyczności (Bremer 2010)². Różnicę między behawioryzmem a funkcjonalizmem przedstawia poniższy schemat:

- wejście → dyspozycje → wyjście – w behawioryzmie same dyspozycje do działania nie są przyczynowo-czynne. Behawioryzm identyfikuje ból z dyspozycją do zachowania się, dzięki czemu teoria ta czyni za-
dość intersubiektywnej komunikowalności, ale o tym, co pojawi się na wyjściu (zachowaniu), decydują w sposób konieczny same wejścia zgodnie ze schematem: bodziec–reakcja;
- wejście → funkcje/role przyczynowe → wyjście – w przypadku funk-
cjonalizmu, stan na wyjściu determinowany jest stanami wewnątrz-
nymi układu określanymi za pomocą funkcji przyczynowych. Moto-
ryczne wyjścia – czyli zachowania się – zależą nie od wejść, ale od
całościowego stanu organizmu.

Funkcjonalistyczna teoria umysłu Putnama krytycznie odnosi się rów-
nież do teorii identyczności. Teoria ta, której najbardziej skrajną wersję
przedstawił J.J.C. Smart, głosi, że zdarzenia mentalne są identyczne ze
zdarzeniami mózgowymi, np. moje doświadczenie bólu jest tym samym
zdarzeniem, co zdarzenie bycia mego mózgu w pewnym stanie neurofizjo-
logicznym M: „nie ma psychofizycznych praw o bikondycjonalnej postaci:
m wtedy i tylko wtedy, gdy f (m – własność mentalna; f – jej fizyczny korelat)”
(Bremer 2010, 126). Zatem te same rodzajowo procesy neurofizjologiczne
odpowiadają za pojawienie się mentalnego zdarzenia bólu zarówno u ludzi,
jak i u innych stworzeń. Nie jest to niemożliwe, ale mało prawdopodobne.
Teoretyk identyczności musi bowiem pokazać, że „każdemu predykatowi
mentalnemu, na przykład odczuwaniu głodu czy dotyku, które można wy-
powiedzieć o tak biologicznie różnych rodzajach (jak na przykład naczelne

² W dużym uproszczeniu można powiedzieć, że behawioryzm logiczny redukuje zdania o przeżyciach (stanach umysłu) do zdań o zachowaniach, natomiast teoria identyczności (inaczej zwana fizykalizmem typów, teorią identyczności rodzajowej, materializmem stanu centralnego czy teorią stanu mózgu) utrzymuje, iż rodzaje przeżyć są tożsame z typami stanów mózgu.

i mięczaki), odpowiada jeden i ten sam korelat fizyczny” (Bremer 2010, 128). Teoretyk identyczności ma kłopot nawet wtedy, gdy weźmiemy pod uwagę osobniki tego samego gatunku, np. ludzi. Indywidualny i niepowtarzalny proces rozwoju i wychowania jednostki oraz odmienność organizacji mózgu każdej jednostki sprawia, że doznanie typu Q u jednostki A nie musi występować z tym samym stanem neurofizjologicznym mózgu, co doznanie tego samego typu Q u jednostki B. Idąc tym tropem myślenia Putnam zauważył, że stany mentalne mają charakter stanów abstrakcyjnych.

Zdaniem Putnama doznawanie bólu (czy innych stanów mentalnych) nie jest stanem mózgu (w sensie stanu fizykalno-chemicznym), ale „funkcjonalnym stanem całego organizmu”, gdzie organizm rozumiany jest jako złożony system materialny. Bremer systematyzuje powyższą hipotezę Putnama w następując sposób (Bremer 2010):

- a) wszystkie organizmy zdolne do odczuwania bólu są komputero-automatami,
- b) każdy organizm jest zdolny do odczuwania bólu, o ile posiada odpowiednią organizację funkcjonalną,
- c) zdolność do odczuwania bólu można przypisać tylko organizmowi jako całości (poszczególne elementy organizmu nie posiadają jako takie odpowiedniej organizacji),
- d) dla każdej funkcjonalnej organizacji istnieje szereg wejść zmysłowych (*sensory inputs*) tego rodzaju, że organizm z daną organizacją jest tylko wtedy w stanie odczuwania bólu, gdy niektóre z jego wejść należą do tego szeregu.

Koniunkcja punktów a–d wyraża tezę o „wielorakiej realizacji”³, która ma być argumentem przeciw teorii identyczności. Samo zaś pojęcie funkcji jest rozumiane w kategoriach przyczynowych: funkcje umysłu realizują procesy przyczynowe.

Główna teza funkcjonalizmu głosi zatem, że te same stany mentalne (ból, procesy poznawcze, np. percepcja, myślenie, wnioskowanie itd.) mogą być zrealizowane przez różne układy fizyczne. Stany mentalne stanowią

³ Putnam stosuje dwie odmiany argumentu z wielorakiej realizacji: empiryczną i pojęciową, zob. (Hensel 2011).

więc funkcję stanów fizycznych. To z kolei oznacza, że identyfikacja danego stanu mentalnego, realizowanego przez system przetwarzający informacje, jest niezależna od fizycznej struktury tego systemu (Kołodziejczyk 2002). Jeśli stany mentalne są przez Putnama rozumiane funkcjonalnie, to na podstawie tezy o wielorakiej realizacji nie można im przyporządkować jednej własności fizycznej. W ten sposób – jak wskazuje Bremer – stany mentalne możemy traktować jako własności drugiego rzędu, które mogą być realizowane przez różne własności pierwszego rzędu (to jest procesy fizyczne) (Bremer 2010).

Koncepcja funkcjonalizmu nabiera charakteru obliczeniowego, gdy postawi się dodatkowy warunek: funkcją umysłu jest przetwarzanie informacji, czyli obliczanie⁴. Putnam posługuje się analogią komputera. Kom-

⁴ Należy tu wspomnieć o tak zwanych modułowych teoriach umysłu. Robert Poczobut wskazuje, że teoria ta oparta jest na trzech podstawowych założeniach: (1) zależność między architekturą mózgu a architekturą umysłu; (2) zaburzenie funkcji poznawczych związane jest zawsze z fizycznym uszkodzeniem mózgu; (3) neuroplastyczność mózgu oznaczająca, że w pełni ukształtowanym mózgu uszkodzone moduły nie zostają zastąpione przez nowe, ale nieuszkodzone części mózgu przejmują ich funkcje. Wśród zwolenników modularnej teorii umysłu samo pojęcie modułu jest różnie rozumiane. I tak Jerry Fodor w systemie umysłu wyróżnił dwa elementy: centralny układ wyższego rzędu związany z pojęciem osoby (odmiennie od Tooby'ego i Cosmidesa), który nie ma specyfiki modularnej, a odpowiedzialny jest za procesy związane z rozumowaniem, podejmowaniem decyzji czy posiadaniem przekonań oraz wyspecjalizowane moduły obliczeniowe. Moduły te automatycznie przetwarzają informacje pochodzące ze środowiska, które pojawiają się „na wejściu” (*input*), układ centralny natomiast operuje na informacjach pojawiających się „na wyjściu” (*output*) poszczególnych modułów. Korelacja między systemem wyższego rzędu a poszczególnymi modułami dokonuje się za pośrednictwem odpowiednich sterowników (*interface*). Nieco inną koncepcję modularności umysłu przedstawia Steven Pinker. Jego zdaniem umysł jest funkcją mózgu. Poszczególne moduły mają charakter funkcjonalny (a nie anatomiczny), są zhierarchizowane i zsynchronizowane w akcie wyższych funkcji poznawczych, przypominając organizm nieredukowalny w prosty sposób do poszczególnych swoich części. Moduły w sposób ewolucyjny wyłoniły się w drodze adaptacji środowiskowej człowieka i mają charakter algorytmiczny (choć Pinker odżegnuje się od komputacjonistycznego porównania, wedle którego umysł ma się do mózgu tak jak *software* do *hardware* komputera). Nie wchodząc w szczegóły, trzeba powiedzieć, że podstawowym problemem generowanym przez modularną koncepcję umysłu jest samo istnienie centralnego układu (Poczobut wskazuje, że samo rozróżnienie na układ centralny oraz moduły peryferyjne jest dyskusyjny i mógłby ewentualnie dotyczyć różnicy ilościowej, a nie jakościowej oraz integrowanie, scalanie obliczeń wykonywanych przez poszczególne moduły, zob. Poczobut 2009, 414–427; M. Hohol 2010, 89–108).

puter jako całość posiada dwa składowe elementy. Posiada bowiem masę, objętość, wewnętrzną strukturę składową pewnych obwodów, posiada cenę itp., ale posiada również pewną własność funkcjonalną, to znaczy program. Program jest własnością niefizyczną, cały bowiem system może tę własność urzeczywistnić niezależnie od swej fizycznej struktury. Odnosząc tę analogię do człowieka, Putnam stwierdza, że własności mentalne, czyli formy myślowe, przedstawienia, doznania, są identyczne z własnościami funkcjonalnymi, są zdarzeniami niefizycznymi o charakterze funkcjonalnym. Tak rozumiane stany mentalne mogą być własnością różnych egzemplarzy różniących się budową fizyczną czy chemiczną. Innymi słowy, stany mentalne rozumiane jako abstrakcyjna funkcja mogą być spełniane niezależnie od swego substratu, niezależnie od swej ontologicznej budowy. Może być nim jakaś istota pozaziemska, jak również stworzony przez człowieka robot. Zdaniem funkcjonalistów jest to najbardziej wiarygodna monistyczna teoria unikająca traktowania umysłu i materii jako dwóch odrębnych rodzajów substancji, czy też dwóch odrębnych jakościowo obszarów własności (Putnam 1995).

Jak wspomnieliśmy wyżej, stanowisko Putnama w kwestii umysłu ewoluowało. Dostrzegł on bowiem, że teoria reprezentowana w *Minds and Machines* generuje szereg problemów związanych z twierdzeniem Gödla⁵. W konsekwencji Putnam odrzuca ontologiczne założenia swojego stanowiska na rzecz interpretacji epistemologicznej (*Representation and Reality* z 1988 r.) i stwierdza, że funkcjonalizm głoszący, iż postawy propozycjonalne są wyłącznie obliczeniowymi stanami, nie może być poprawny. Jak pisze Kołodziejczyk, „mając na uwadze trudności funkcjonalizmu skrajnego oraz odrzucenie tego stanowiska przez samego autora *Representation...*, zdziwienie może budzić fakt, iż skrajny funkcjonalizm stał się dominującym kierunkiem w badaniach nad sztuczną inteligencją” (Kołodziejczyk 2002, 8).

⁵ Ponadto ujęcie zaproponowane w ramach funkcjonalizmu skrajnego nie wyjaśnia pytań związanych z: problemem postaw propozycjonalnych, semantyczną interpretacją procedur obliczeniowych, czy intencjonalnym odniesieniem stanów mentalnych danego systemu, zob. (Kołodziejczyk 2002).

Stanowisko funkcjonalistyczne jest bowiem powszechnie przyjmowane przez zwolenników tak zwanej silnej wersji sztucznej inteligencji. Wychodzą oni z założenia, że skoro stany mentalne dają się zredukować do stanów funkcjonalnych i mogą być spełniane niezależnie od budowy fizycznej swego podłoża, to każdy system fizyczny odpowiednio zaprogramowany, będzie obdarzony umysłem spełniającym dokładnie takie same funkcje jak umysł człowieka. Jak nieco ironicznie pisze o tym Searle: „Zatem, jeśli zrobilibyśmy komputer ze starych puszek po piwie, napędzanych wiatrakami, jeśli zaprogramowalibyśmy go odpowiednio, byłby on obdarzony umysłem” (Searle 1995, 76). Problem, zdaniem Searle’a, nie leży w tym, że taki system *mógłby* myśleć i odczuwać, ale w tym, że *musiałby* on myśleć i czuć, jedynym bowiem warunkiem myślenia i odczuwania jest zainstalowanie odpowiedniego programu. Umysł jest tożsamy z programem cyfrowego komputera, zatem odpowiednio rozwinięty i skomplikowany program wstawiony do maszyny musi wytworzyć świadomość. Przyjmujący teorię sztucznej inteligencji twierdzą bowiem, że proces myślenia może być w pełni opisany przez algorytmy: „Myślenie polega na obliczeniach, a w szczególności świadome doznania powstają wskutek odpowiedniego procesu obliczeniowego” (Penrose 2000, 31).

Czy jednak myślenie ludzkie da się sprowadzić jedynie do czynności obliczeniowych i przetwarzania?

2. Myślenie ludzkie a przetwarzanie informacji

John Searle, przeprowadzając eksperyment myślowy zwany Chińskim pokojem, wykazuje, że samo algorytmiczno-syntaktyczne manipulowanie danymi dostarczonymi z zewnątrz nie wystarcza do wytworzenia stanów mentalnych, czy szerzej, do wytworzenia świadomości⁶. Eksperyment przebiega tak: wyobraźmy sobie, że jestem zamknięty w pokoju, w którym znajduje się mnóstwo koszy wypełnionych znakami w języku chińskim.

⁶ Ten argument Johna Searle’a pojawił się jako sprzeciw profesora z Berkeley wobec mocnego programu sztucznej inteligencji po raz pierwszy w 1980 roku i częściowo zmieniał swoje brzmienie, zob. (Kloch 1995, 75–101).

Nie znam chińskiego, ale mam ze sobą książkę napisaną w moim ojczystym języku, która zawiera reguły manipulowania w sposób syntaktyczny, formalny znakami języka chińskiego. Reguły te mają taką na przykład treść: „Wybierz ten znak z podwójnym zakrętasem z kosza numer jeden i połącz go za znakiem z dwoma zawijasami z kosza numer dwa”. W pokoju pojawiają się podawane przez kogoś z zewnątrz nowe znaki chińskie, ja zaś otrzymuję instrukcje, jakimi kombinacjami znaków budowanych za pomocą reguł z mojej książki mam odpowiadać na te znaki pojawiające się z zewnątrz. Załóżmy, że ja, siedząc w pokoju, nie wiem, że przysyłane z zewnątrz kombinacje znaków są pytaniami, a moje wysyłane kombinacje są z kolei odpowiedziami. Załóżmy również, że nabrałem takiej biegłości w tworzeniu kombinacji znaków, że nie są one odróżnialne od kombinacji znaków tworzonych przez kogoś, kto rzeczywiście zna język chiński.

Na podstawie tego eksperymentu Searle wykazuje, że błędne są poglądy zwolenników sztucznej inteligencji. Przedstawiając wnioski Searle’a będę nadal posługiwał się formą pierwszoosobową: przebywając w pokoju i nie znając języka chińskiego, nie rozumiem ani słowa z przekazywanych mi z zewnątrz pytań, nie rozumiem też ani słowa z dawanych przeze mnie odpowiedzi. Posługuję się tylko symbolami na poziomie syntaktycznym, a więc pozbawionym dla mnie jakiegokolwiek sensu. Interpretacji tego, co robię, dokonuje dopiero osoba na zewnątrz, która wprowadza znaki do pokoju i odbiera te, które ja oddaję (osoba ta, jeśli nie ma pełnej wiedzy na temat mojej znajomości języka chińskiego, może być przekonana na podstawie wysyłanych przeze mnie znaków, że doskonale znam chiński. Osąd tej osoby będzie oczywiście błędny). Analogicznie pracuje procesor cyfrowego komputera. Przyjmuje przesyłane mu z zewnątrz dane, przetwarza je i podaje na zewnątrz. Sama manipulacja symbolami nie wystarcza do uzyskania świadomości. Przebywając w pokoju i nie znając języka chińskiego, działając analogicznie jak procesor komputera, nie mogę przejść z poziomu syntaktyki do semantyki, nie jestem świadomy treści ani pytań, ani dawanych przeze mnie odpowiedzi. Podobnie program komputerowy nie jest świadomy przekazywanych mu z zewnątrz informacji, czyli „danych”, nie odróżnia ich od siebie i nie odnosi do niczego. Nie jest on w stanie przejść

z poziomu syntaktyki do semantyki, gdy tymczasem każdy stan umysłowy – jak twierdzi Searle – jest stanem aktualnie bądź potencjalnie świadomym (Miłkowski, Poczobut 2012, 17).

Podobny do searlowskiego jest eksperyment myślowy Neda Blocka, znany pod nazwą „Wielki Mózg Chin”. W eksperymencie tym Chiny zamienione są w symulację ogromnych rozmiarów mózgu, a każdy mieszkaniec Chin jest jednym neuronem tego mózgu. Neurony otrzymują z satelity instrukcje i wykonując je, przesyłają między sobą sygnały za pomocą latarek. Zgodnie ze stanowiskiem funkcjonalizmu ogromny mózg jest modelem mózgu człowieka. Czy przez sam fakt manipulacji sygnałami bez znajomości ich znaczenia Mózg Chin będzie miał stany umysłowe? Czy wytworzy świadomość?

Kontrargument Searle’a i Blocka przeciwko silnej wersji sztucznej inteligencji stawiającej znak równości między ludzkim umysłem a komputerem zwraca uwagę na zasadniczą, nieprzekraczalną różnicę, jaka istnieje między nimi. Program komputerowy i przeprowadzane przez niego operacje można scharakteryzować czysto formalnie i opisać za pomocą abstrakcyjnych symboli, na przykład sekwencji zer i jedynek. Symbole te jednak nie mają żadnego znaczenia semantycznego, nie odnoszą się do niczego konkretnego. I w tym tkwi zasadnicza różnica niepozwalająca na utożsamienie myśli czy procesu umysłowego z procesem wykonywanym przez program komputerowy. Proces umysłowy bowiem jest zawsze wyposażony w jakąś treść, przy czym nie przeczy się oczywiście temu, że jest również jakimś przetwarzaniem informacji. Myśl nie jest jednak tylko ciągiem abstrakcyjnych symboli, jest zawsze czymś więcej, jest zawsze myślą o czymś, ma intencjonalne odniesienie. Searle nie zaprzecza, że w ludzkim umyśle następuje przetwarzanie informacji, postuluje jednak, aby w tym kontekście „informacje” rozumieć w sensie semantycznym. Umysł ludzki odnosi się do znaczenia, a przetwarzanie informacji w takim rozumieniu wymaga stanów intencjonalnych. Zbiór symboli programu komputerowego nie odnosi się do niczego, pozostaje na poziomie syntaktycznym, gdy tymczasem myśl wyposażona w treść posiada semantykę, ma jakieś znaczenie.

Zredukowanie myślenia tylko do procesu obliczeniowego o charakterze algorytmicznym, jak to czynią zwolennicy silnej wersji sztucznej in-

teligencji, generuje szereg trudności teoretycznych. Algorytm bowiem to „dokładny opis postępowania gwarantujący osiągnięcie określonego rezultatu w skończonej ilości elementarnych kroków” (Kulicki 2000, 188). Najistotniejsze cechy algorytmu to zatem niezmienna sekwencyjność kolejno podejmowanych kroków oraz jednoznaczność determinująca to, że każdorazowe wykonanie poszczególnej czynności elementarnej odbywa się zawsze w taki sam sposób. Jeśli procesy poznawcze człowieka i maszyny mają charakter obliczeniowy – to procesy te można by wyrazić za pomocą skończonego ciągu symboli powiązanych ze sobą za pomocą reguł transformacji zapodmiotowanych w strukturze systemu przetwarzającego informacje (Kołodziejczyk 2002). Prowadzi to do redukcjonistycznego programu opisu procesów poznawczych w kategoriach logiki formalnej, natomiast zachodzenie procesu poznawczego uwarunkowane jest realizacją określonego algorytmu.

Samo porównanie umysłu ludzkiego z maszyną w zakresie przeprowadzania wnioskowania ukazuje niemożliwość ich utożsamienia pod względem dokonywania operacji poznawczych. Zaprogramowana bowiem maszyna nie może wykroczyć poza formalny system, czyli nie może działać intuicyjnie, jest ona bowiem zdeterminowana przez zbiór reguł, na podstawie których została zaprogramowana, a którego sama z siebie nie jest w stanie przekroczyć. Żaden program komputerowy nie posiada zdolności dowodzenia prawdziwości zdań samozwrotnych, czyli wykraczających poza system (Kołodziejczyk 2002).

Nasze codzienne doświadczenie wskazuje, że myślenie nie jest wyłącznie mechaniczne, często nie możemy wskazać kolejnych jego kroków, a czasami po prostu ich nie ma, myśl bowiem pojawia się nagle. „Nasze procesy mentalne, które są podstawą naszego JA, naszego życia, są nie tylko abstrahowaniem i mechanicznym sortowaniem, lecz czymś podmiotowym, osobistym i jako takie obejmują nie tylko klasyfikowanie i dzielenie na kategorie, ale też ciągłe odczuwanie i poddawanie osądowi” (Sacks 1985, 41). Można przyjąć, że myśl zawiera cechy algorytmiczne, gdy myślimy np. o skoku z dachu, obliczając jego wysokość i ewentualne szkody, jakie możemy ponieść. Odwołanie jednak do algorytmu nie jest wystarczającym

wyjaśnieniem tego, jak przebiega proces myślowy przy tworzeniu na przykład skomplikowanego utworu muzycznego. Tym bardziej że utwór muzyczny nie jest tylko jakąś kombinacją następujących po sobie dźwięków. Mówiąc językiem nieco poetyckim, w układzie dźwięków i budowanych z nich taktów kryje się jakiś zamysł, jakaś myśl autora, który w ten sposób wyraża swoje emocje. Tworzone przez niego dzieło jest wyrazem buntu lub akceptacji wobec określonej wizji świata⁷. Wystarczy pójść kilka razy na dobry koncert jazzowy, by przekonać się, że muzyka grana przez artystę jest czymś daleko więcej niż tylko odtworzeniem zapisu nutowego. Ten ostatni stanowi bowiem jedynie grunt do niewyobrażalnej ilości wariacji na podstawowy temat muzyczny, a końcowy efekt – dzieło, którego można wysłuchać – jest wypadkową nieuchwytną dla jakiegokolwiek programu komputerowego „gry” emocji, doznań, nastrojów zarówno po stronie artysty, jak i słuchacza. Same schematy, programy, algorytmy są niewystarczające, nie mogą nas bowiem obdarzyć „wizyjną, dramatyczną i muzyczną

⁷ Ewolucja poglądów Putnama zdaje się potwierdzać te intuicje. W kwestii relacji psychofizycznej głównymi rozwiązaniami wśród filozofów były wersje fizykalizmu o charakterze redukcjonistycznym. Pojawienie się monizmu anomalnego sprawia odwrót od redukcjonizmu, ale jednocześnie nie osłabiło popularności fizykalizmu. Należało więc stworzyć taką teorię fizykalizmu, która odżegnuje się od redukcjonowania tego, co mentalne, do tego, co fizyczne. Fizykalizmem tego typu jest *superweniencja mereologiczna*: systemy o takich samych mikrostrukturach mają takie same makrowłasności systemowe. Brzozek podaje taki przykład: „Znakomita większość dzisiejszych muzyków korzysta mniej lub bardziej z pomocy programów komputerowych, umożliwiających obróbkę materiału muzycznego, od drobnej kosmetyki po tworzenie gotowych, kompletnych utworów. Wyobraźmy sobie taki program, nieskomplikowany, z którego korzysta właśnie dwóch muzyków. Wypełnili oni swoje pięciolinie w dokładnie ten sam sposób: wykorzystali te same nuty, pauzy, akcenty, rozmieszczając je w tych samych miejscach. Otrzymali więc dwie nieodróżnialne partytury. Czy oba utwory będą identyczne w brzmieniu? Zakładając idealne wykonanie obu tych utworów zgodnie z partyturą, nie będziemy w stanie odróżnić ich od siebie – będzie nam się wydawało, iż słyszymy ten sam utwór grany jeden po drugim. Identyczność pod względem zapisu muzycznego pociąga identyczność w brzmieniu. Zmieniając zaś miejscami kilka nut, otrzymamy różnicę w brzmieniu. [...] zapis nutowy dwóch utworów jest własnością ze zbioru B, to zaś, [...] co słyszymy podczas wykonywania utworów (oraz to, co odczuwamy), to własności ze zbioru A. Nieodróżnialność własności ze zbioru B pociąga nieodróżnialność własności ze zbioru A”. Jednak zdaniem Putnama – jak interpretuje jego tekst *Mózgi w naczyniu* Robert Poczobut – psychosemantyka nie mikroredukuje się do neurobiologii, ponieważ treści umysłowych nie determinują wyłącznie wewnętrzne, neurobiologiczne stany systemu poznawczego; zob. (Brzostek 2011, 215–224).

wartością przeżycia – taką żywą, osobistą wartością, która sprawia, że jest to właśnie przeżycie” (Sacks 1995, 197).

Myśl jest zawsze myślą o czymś, ma treść i intencjonalne odniesienie, czego nie można powiedzieć o komputerowym programie. Kryje się w tym jednak jeszcze jedno bardzo istotne założenie: myśl jest też zawsze myślą kogoś. Nawet, jeśli przyjmiemy założenie, że myślenie jest tylko przetwarzaniem informacji, bodźców pochodzących z zewnątrz (z powyższych rozważań wynika, że w przypadku ludzkiego umysłu te informacje są przeniknięte semantyczną treścią), to te informacje są odniesione do jakiegoś świadomego podmiotu, są świadomie prezentowane sobie przez jakieś „ja”. Świadomość owych wrażeń, informacji, a więc świadomość intencjonalna zakłada przedrefleksyjną, bezpośrednią świadomość siebie (Frank 2002). Jest ona bezpośrednia w tym sensie, że znajomości własnego życia psychicznego nie uzyskuje się przez akty sądzienia lub postrzegania. Pojęcia takie jak „ból zęba” czy „bycie zakochanym” są nazwami tego, co jest przez podmiot doznawane. Świadomość intencjonalną, świadomość czegoś trzeba odróżnić od przedrefleksyjnej świadomości siebie. Różnica ta jednak nie świadczy o tym, że świadomość czegoś jest obiektem prezentowanym świadomości przedrefleksyjnej, obie one bowiem tworzą funkcjonalną jedność. Istotą tej tezy, bronionej z różnych pozycji przez J.G. Fichtego oraz F. Brentano, dzisiaj przez H. Castaneda czy M. Franka, jest absolutna nieredukowalność bezpośredniej świadomości siebie. Żadnemu systemowi komputerowemu, choćby najbardziej zaawansowanemu na miarę możliwości naszego wyobrażenia, nie przypiszemy tego specyficznego stanu epistemicznego wyprzedzającego wszelką refleksję i umożliwiającego jej wyjaśnienie bez popadania w błędne koło.

Funkcjonalizm przyjmuje, że ten sam typ zdarzeń mentalnych może być różnie realizowany przez całkiem inny typ zdarzeń neurofizjologicznych u różnych osób w różnym czasie. Ale też jednemu typowi zdarzeń mentalnych u osobnika A nie musi odpowiadać dokładnie jeden typ zdarzeń neurofizjologicznych. Teoria ta jednak ostatecznie zdaje się zmierzać do tego, że w pełni rozwinięta neuronauka zastąpi pojęcia mówiące

o stanach mentalnych pojęciami i prawami fizycznymi⁸. Zamiast wtedy mówić: „wzrost aktywności wulkanicznej na świecie jest realnym dla niego zagrożeniem”, będę mówił: „mój mózg jest w stanie pobudzenia włókien nerwowych typu C o natężeniu K w czasie t”. Tak optymistyczna wizja rozwoju neuronauki jest jednak przesadną wiarą w eksplanacyjną moc nauk szczegółowych. Wydaje się, że próba znalezienia zmian fizycznych mózgu identycznych (czy odpowiadających) ze stanami mentalnymi typu ból, zakochanie, przekonania itp. jest analogiczna do próby znalezienia najmniejszego punktu czy najmniejszej wartości w przedziale od 0 do 1. Przecież na indywidualny sposób przeżywania świata, doznawania stanów mentalnych, ich natężenia, sposobu reagowania na nie itd. ma wpływ przeogromna i jak się wydaje, nieprzeliczalna ilość czynników. Wiele z nich jest też w ogóle nieuświadomionych. Wszystko to wpływa na odmiennność i niepowtarzalność organizacji mózgu każdej jednostki ludzkiej. Doświadczenie pokazuje również, że te same czynniki zewnętrzne, okoliczności i doznania, doświadczane w różnym czasie, wywołują różne stany mentalne. Czy są one identyczne ze zmianami fizycznymi mózgu? O jaką identyczność chodzi? Opierając się na definicji identyczności zaproponowanej przez Leibniza mówiącej, że dwa obiekty, przedmioty itp. są identyczne, gdy przysługują im te same własności i będąc konsekwentny w przyjmowaniu teorii identyczności, trzeba by powiedzieć, że mojemu stanowi neuronalnemu identycznemu z przekonaniem „wzmóżonej aktywności wulkanicznej” należy przypisać własność lęku. A to przecież jest absurdalne. Co prawda, Putnam wyróżnia dwa rozumienia pojęcia własności: własność jako predykat oraz pojęcie własności rozumiane jako syntetyczna identyczność własności: opis tego samego ciała za pomocą własności temperatury *x ma taką a taką temperaturę* jest tą samą syntetycznie własnością, co posiadanie przez to

⁸ Teoria identyczności przyjmuje tezę redukcji interdyscyplinarnej – zredukowanie teorii posługującej się językiem mentalistycznym do teorii posługującej się językiem fizykalistycznym. Może ona przyjąć dwie formy: redukcję miękką, eksplanacyjną – rozwój nauki prowadzi do tego, że procesy mentalne będą wyjaśniane za pomocą języka fizykalistycznego (stanowisko to nazywane jest przez W. Quine’a materializmem nieredukującym, odwołuje się również do teorii zwanej monizmem anomalnym, głoszącej, że nigdy się nie da sformułować wszystkich praw mentalnych za pomocą praw fizycznych) oraz redukcja twarda, eliminacyjna – zastępuje byty mentalne bytami fizycznymi, zob. Judycki 2004).

ciało pewnej energii *x ma „s” energię kinetyczną*. To rozróżnienie jednak nie wyjaśnia, czy własności przypisywane przez nas przekonaniu są identyczne z własnościami fizycznymi mózgu. Jeśli moje przekonania są identyczne ze stanami neuronalnymi w moim mózgu, to czy zmiany neuronalne, zmiany fizyczne spowodowane na przykład wylewem krwi do mózgu spowodują wytworzenie identycznego z nimi przekonania? Jak się wydaje, te dwie kategorie jakości są do siebie niesprowadzalne.

Według M. Minsky’ego, jeśli zrozumiemy dla człowieka, tak jak dla maszyny, strukturę i program, to zniknie uczucie tajemniczości i samouwielbienia (Minsky 1972). Myślę, że w świetle powyższych rozważań niesprowadzalność ludzkiego myślenia wyłącznie do przetwarzania informacji, wbrew przekonaniom Mińskiego i innych różnego rodzaju redukcjonistów, jest możliwa do obronienia. Teoria funkcjonalizmu nie wyjaśnia w sposób zadowalający, czym jest umysł. Często odwołujemy się do definiowania czegoś przez wskazanie na spełniane funkcje. Tak definiujemy gaźnik czy pułapkę na myszy. Ale czy wyjaśnia nam to ostatecznie, czym jest gaźnik? Funkcjonalizm, utożsamiając stany mentalne z własnościami funkcjonalnymi, nie daje satysfakcjonującego wyjaśnienia, czym w swej istocie jest umysł, czym są stany mentalne, w jaki sposób powstają. Myślenie człowieka nie jest tylko syntaktycznym przetwarzaniem informacji, nie wyjaśnia go odwołanie do funkcji algorytmicznych. Myśl zawsze wyposażona jest w treść i intencjonalne, semantyczne odniesienie, czego nie da się powiedzieć o najbardziej nawet wyrafinowanych technicznie programach komputerowych. Absolutnie nieredukowalna jest przedreflexyjna, bezpośrednia świadomość samego siebie. W końcu przekonania, jakie żywimy, nasze stany mentalne nie dadzą się sprowadzić do identycznych z nimi stanów neurofizycznych mózgu.

Bibliografia

- Bremer, J. 2010. *Wprowadzenie do filozofii umysłu*. Kraków: WAM.
 Brzostek, B. 2011. „Superweniencja – pytanie o trywialność.” *AVANT. Pismo Awangardy Filozoficzno-Naukowej* II 2: 215–224.

- Frank, M. 2002. *Świadomość siebie i poznanie siebie*. Warszawa: Oficyna Naukowa.
- Heil, J. 2001. „Funkcjonalizm, realizm i poziomy bytu.” W *Pragmatyzm i filozofia Hilarego Putmana*, red. U. Żegleń, 129–147. Toruń: Wydawnictwo UMK.
- Hensel, W.M. 2011. „Dwa funkcjonalizmy Hilarego Putnama, czyli kawałek historii z morałem.” *Diametros* N 29 (September).
- Hohol, M. 2010. „Umysł: system sprzeczny, ale nie trywialny.” *Zagadnienia filozoficzne w nauce* XLVII: 89–108.
- Judycki, S. 2004. *Świadomość i pamięć*. Lublin: Towarzystwo Naukowe KUL.
- Kloch, J. 1995. „Chiński pokój. Eksperyment myślowy Johna Searle’a.” *Zagadnienia filozoficzne w nauce* XVII: 75–101.
- Kołodziejczyk, P. 2002. „Funkcjonalizm jako filozoficzna podstawa teorii Sztucznej Inteligencji.” Dostęp: 4 VII. <http://www.kognitywistyka.net>.
- Kulicki, P. 2000. „Algorytm.” W *Powszechna encyklopedia filozofii*. T. 1, red. A. Maryniarczyk i inni, 188–189. Lublin: Polskie Towarzystwo Tomasza z Akwinu.
- Minsky, M. 1972. „Na drodze do stworzenia sztucznej inteligencji.” W *Maszyny matematyczne i myślenie*, 378–424, red. E. Feigenbaum, J. Feldman, Warszawa.
- Penrose, R. 2000. *Cienie umysłu. Poszukiwanie naukowej teorii świadomości*, Poznań: Zysk i S-ka.
- Poczobut, R. 2007. „Eksternalizm treści umysłowych a superweniencja.” *Kognitywistyka i Media w Edukacji* 9: 82–106.
- . 2009. *Między redukcją a emergencją. Spór o miejsce umysłu w świecie fizycznym*, Wrocław: Wyd. Uniwersytetu Wrocławskiego.
- Miłkowski, M., Poczobut, R., red. 2012. *Przewodnik po filozofii umysłu*. Kraków: WAM.
- Putnam, H. 1995. „Umysł a ciało.” W *Filozofia umysłu*, tłum. B. Chwedeńczuk: 185–213. Warszawa: Wydawnictwo SPACJA.
- Sacks, O. 1985. *Mężczyzna, który pomylił swoją żonę z kapeluszem*, przekł. B. Lindenberg. Poznań: Z-sk i Spółka.
- Searle, J. 1995. *Umysł, mózg i nauka*. Warszawa: PWN.